

文章编号: 1002-1566(2019)04-0661-14  
DOI: 10.13860/j.cnki.sltj.20190528-003

# 贝叶斯共同成分模型拓展及其在高血压风险评估的应用

徐丽<sup>1</sup> 方亚<sup>2</sup>

(1. 广东外语外贸大学数学与统计学院, 广东 广州 510420; 2. 厦门大学公共卫生学院, 福建 厦门 361102)

**摘要:** 尝试引入贝叶斯样条对贝叶斯共同成分模型进行拓展, 并基于 1991-2011 年《中国健康与营养调查》的实际数据探究高血压风险的空间变异与时间变异。研究发现: 高血压风险 (尤其是男性) 大致呈现出北方高于南方的空间格局, 且这一空间分布模式具有相对稳定性, 这可能与北方各省份相似且稳定的生活方式 (饮酒、膳食高盐等) 有关。拓展后的 SCM 模型更充分地挖掘样本信息, 揭示出高血压风险的空间格局及其随时间演变的规律, 克服了空间分析时样本量相对不足的局限。

**关键词:** 共同成分模型; 贝叶斯样条; 相对风险; 高血压

**中图分类号:** C81, O212

**文献标识码:** A

## Bayesian Shared Component Model Extension and Its Application in Risk Assessment of Hypertension

XU Li<sup>1</sup> FANG Ya<sup>2</sup>

(1. College of Mathematics and Statistics, Guangdong University of Foreign Studies, Guangdong Guangzhou 510420, China, 2. School of Public Health, Xiamen University, Fujian Xiamen 361102, China)

**Abstract:** This paper tried to extend the existing Shared Component Model (SCM) by introducing the Bayesian spline and employed the extended model on the basis of 1991-2011 China Health and Nutrition Survey (CHNS). Results showed that the risk of hypertension (especially for males) generally showed a spatial pattern that is higher in the Northern China than that in the Southern China. Also, this pattern was relatively stable during the study period. This pattern may be related to the similar and stable lifestyle in the Northern areas, such as drinking alcohol, high salt diet, etc. The extended SCM is able to explore the sample information fully and reveal the spatial pattern and trend over time with respect of the risk of hypertension, thus overcome the limitation of limited sample size of spatial analysis.

**Key words:** shared component model, Bayesian spline, relative risk, hypertension

## 0 引言

高血压是最常见的慢性病, 也是心脑血管疾病最主要的危险因素。根据世界卫生组织 (WHO) 2013 年的报告, 全球每年有 1700 万人死于心血管疾病, 占死亡总数的 1/3, 其中

收稿日期: 2016 年 3 月 20 日

收修改稿日期: 2018 年 6 月 23 日

基金项目: 2015 广东省教育厅创新强校青年创新人才类项目 (自然科学类) (2015KQNCX036)。

有 940 万死于高血压的并发症。世界各国尤其是低收入和中等收入国家居高不下的高血压患病率严重消耗了医疗和社会资源,给家庭和国家造成了沉重的负担<sup>[1]</sup>。据 WHO 报告预测,在 2011 年至 2025 年期间,低收入和中等收入国家与非传染性疾病相关的累计产出损失预计将达到 7.28 万亿美元,而年度产出损失共计约为这些国家国内生产总值的 4%,其中心血管疾病(包含高血压)约占了一半的成本<sup>[2]</sup>。

国内外研究表明,高血压患病率的日益升高与超重或肥胖、膳食高盐、吸烟、中度或重度饮酒、睡眠时间缩短等因素存在着紧密联系<sup>[3-8]</sup>。而国外某些研究还发现,这些因素对高血压的作用机制还可能与性别有关<sup>[9-12]</sup>。综合已有研究发现,国内外学者对影响高血压发生的危险因素的相关研究成果颇丰,但也存在以下不足:①从研究视角来看,现有研究大多基于个体层面而鲜少从区域层面研究高血压风险及其潜在危险因素;②从研究方法来看,现有研究主要采用 logistic 回归、Meta 分析、Cox 比例风险回归等传统方法。这些传统方法均没有利用样本中所包含的时间与空间信息,且其大多以大样本为前提,具有一定局限性。

为此,本文拟利用贝叶斯共同成分模型(Shared Component Model, SCM)在样本量较小时仍然能够获得直观、稳健的相对风险估计的特点,分别研究男性与女性高血压风险的空间分布模式及其随时间变化的趋势,探讨高血压风险区域差异或性别差异形成的可能原因,从而为卫生部门有效防控高血压提供实证依据与决策参考。

## 1 贝叶斯共同成分模型(SCM)研究概述

SCM 是基于分层贝叶斯模型(Hierarchical Bayesian Model, HB)框架对多疾病进行风险估计的方法。SCM 最早由 Knorr 和 Best (2001)<sup>[13]</sup>提出,用于两种疾病的联合分析。同年, Dabney 和 Wakefield 探讨了将协变量加入两种不同疾病时的建模方式,并与单种疾病的模型进行了比较<sup>[14]</sup>,随后 Held 等 (2005)<sup>[15]</sup>将其拓展到四种疾病的情形。SCM 模型最初仅包含空间信息,只能用于展示疾病风险的空间变异。

随着用于执行贝叶斯模型的 Openbugs 等免费软件的普及,国内外研究者尝试同时将空间与时间信息融入分层贝叶斯模型,构建多种贝叶斯时空模型<sup>[16-21]</sup>。其中, Richardson (2006)<sup>[16]</sup>最早尝试构建 SCM 时空模型;随后,一些研究者将其拓展为多种疾病的时空分析(Tzala 和 Best, 2008<sup>[17]</sup>; Baker 等, 2017<sup>[18]</sup>; Mahaki 等, 2017<sup>[19]</sup>)。与 Richardson (2006)<sup>[16]</sup>类似, Susanna (2015)<sup>[21]</sup>也是以性别为研究视角,探讨男性与女性肺癌的共同危险因素及其随时间演变的规律,但对 Richardson (2006)<sup>[16]</sup>的模型进行了修正,使其能够拟合数据特别稀疏情形下的样本变异。

已有 SCM 时空模型的特点是:主要采用参数的方法拟合时间趋势,且以罕见疾病(如癌症)为研究对象。而高血压为非罕见疾病,其患病特点与癌症等罕见疾病有较大的不同。因此,为充分利用样本信息,挖掘高血压患病风险的空间分布模式及其随时间发生的演变,本文拟尝试将贝叶斯 B 样条引入 SCM 模型进行拓展,丰富 SCM 模型的理论及运用研究。

## 2 贝叶斯共同成分模型(SCM)形式及其时空拓展

### 2.1 贝叶斯共同成分(SCM)空间模型

SCM 模型基本形式如下:

$$O_{ji} \sim \text{bin}(n_{ji}, p_{ji}), \quad (1)$$

$$\log \text{it}(p_{ji}) = \alpha_j + \varepsilon_{ji}, \quad (2)$$

$$\varepsilon_{1i} = \varphi_i \delta + v_{1i}, \quad (3)$$

$$\varepsilon_{2i} = \varphi_i / \delta + \beta_i + v_{2i}, \quad (4)$$

$$\eta_1 = \text{var}(\varphi_i \delta) / (\text{var}(\varphi_i \delta) + \text{var}(v_{1i})), \quad (5)$$

$$\eta_2 = \text{var}(\varphi_i / \delta) / (\text{var}(\varphi_i / \delta) + \text{var}(\beta_i) + \text{var}(v_{2i})), \quad (6)$$

其中,  $i$  表示区域,  $j$  表示性别 (1 代表男性, 2 代表女性)。  $O_{ji}, n_{ji}, p_{ji}$  分别表示某区域上男性 (或女性) 高血压的患病人数、区域总观测人数、患病率。  $\alpha_j$  表示固定效应, 为便于与疾病相对风险的量级相对应, 通常将其表示为相对风险的形式, 如 RR。  $\alpha_j = \exp(\alpha_j)$ ;  $\varepsilon_{ji}$  表示随机效应, 根据 SCM 模型基本思想分解为男性与女性共同的随机变异  $\varphi_i$  (权重分别为  $\delta$  与  $1/\delta$ ) 以及与性别相关的随机变异  $v_{ji}$ , 而  $\beta_i$  则表示剔除协变量效应后, 男性与女性高血压风险性别差异的空间变异。  $\eta_j$  表示男性或女性空间总体随机效应变异中由共同成分贡献的比例, 其数值大小可以反映出男性与女性共同成分变异是否显著存在。

需要注意的是: ①已有 SCM 模型相关研究主要关注癌症等罕见疾病, 因此采用泊松分布建模, 而本研究中高血压为常见疾病, 因此改用二项分布建模; ②  $\varphi_i$  的权重设为  $\delta$  与  $1/\delta$  满足了权重对数之和为 0 的条件, 确保模型具有可识别性; ③为量化区域层面上自变量与结果变量之间的作用程度, 可在式 (2) 中加入区域协变量  $x_{ji}$ , 并用  $\beta_j$  表示其平均效应。

共同成分  $\varphi_i$  与性别特有成分  $v_{ji}$  采用卷积先验设定, 分解为空间非结构化与结构化随机效应, 即:

$$\varphi_i = ush_i + ssh_i, \quad (7)$$

$$v_{ji} = bind_{ji} + bspat_{ji}, \quad (8)$$

其中,  $ush$ 、 $ssh$  与  $bind$ 、 $bspat$  分别表示共同成分与疾病特有成分的空间非结构化与结构化随机效应。

## 2.2 贝叶斯共同成分 (SCM) 时空拓展

### 2.2.1 时空模型构建

本文尝试在借鉴以往研究的基础上对 SCM 模型进行拓展, 探讨样条法在多疾病情形下拟合时间趋势中的作用, 丰富 SCM 模型理论与应用研究, 为多疾病同时进行监测提供方法论指导。

引入贝叶斯样条的 SCM 时空模型构建如下:

$$O_{jit} \sim \text{bin}(n_{jit}, p_{jit}), \quad (9)$$

$$\log \text{it}(p_{jit}) = \alpha_j + \varepsilon_{jit}, \quad (10)$$

$$\varepsilon_{1it} = (b_{0i} + RS_{0i}(t))\delta_t + S_1(t) + b_{1i} + RS_{1i}(t) + \beta_{it}, \quad (11)$$

$$\varepsilon_{2it} = (b_{0i} + RS_{0i}(t))/\delta_t + S_2(t) + b_{2i} + RS_{2i}(t) + \beta_{it}, \quad (12)$$

$$\eta_1(t) = \text{var}((b_{0i} + RS_{0i}(t))\delta_t) / \text{var}(\varepsilon_{1it}), \quad (13)$$

$$\eta_2(t) = \text{var}((b_{0i} + RS_{0i}(t))\delta_t^{-1}) / \text{var}(\varepsilon_{2it}), \quad (14)$$

其中,  $i$  表示区域 ( $i = 1, 2, \dots, N$ ),  $j$  表示性别 (1 代表男性, 2 代表女性),  $t$  表示年份 ( $t_1, \dots, t_T$ ), 并以 2000 年为中心;  $b_{0i}$  为男性与女性共同空间变异 (空间主效应), 而  $RS_{0i}(t)$  为

男性与女性共同的时空变异(时空主效应),从而  $(b_{0i} + RS_{0i}(t))$  表示男性与女性共同的总体变异;  $S_j(t)$  表示性别特有的时间变异(性别-时间交互效应),而  $b_{ji}$  表示性别特有的空间变异(性别-空间交互效应),  $RS_{ji}(t)$  则表示性别特有的时空变异(性别-时空交互效应),  $\beta_{it}$  为性别差异的时空变异(时空交互效应);  $\delta_t$  与  $1/\delta_t$  表示各年份男性或女性共同成分  $(b_{0i} + RS_{0i}(t))$  的权重,  $\eta_j(t)$  表示各年份男性或女性空间总体随机效应变异中由共同成分贡献的比例;其余变量的含义与 2.1 节 SCM 模型基本相同,只是将二维拓展为三维。

与 2.1 节类似,这里将男性与女性共同成分权重设为  $\delta_t$  与  $1/\delta_t$  满足了权重对数之和为 0 的条件,确保模型具有可识别性;而式 (13) 与式 (14) 分别暗含着式 (11) 与式 (12) 中各成分相互独立的假定。而且,为保持模型的可识别性,式 (11) 与式 (12) 中的某些成分会根据模型的运行结果进行适当调整。

### 2.2.2 模型诊断与模型评估

SCM 模型的贝叶斯推断借助马尔科夫链蒙特卡洛模拟 (Markov Chain Monte Carlo, MCMC) 法完成。模型收敛性诊断采用经典的方差比值法,即 Brooks-Gelman-Rubin 统计量<sup>[22]</sup>并结合动态轨迹图 (trace) 与自相关图 (Autoregressive Correlation Function, ACF) 等图形综合诊断。模型评估采用离差信息准则 (Deviance Information Criterion, DIC, Spiegelhalter, 2002)<sup>[22]</sup>,一般对于同一数据而言,模型的 DIC 越小,说明模型更优。

### 2.2.3 样条函数及结点选择

这里,  $RS_{0i}(t)$ 、 $S_j(t)$  与  $RS_{ji}(t)$  均设定为样条平滑函数,并假定回归立方 B 样条为基函数,结点数为  $L$ 。需要说明的是,尽管 B 样条、P 样条与平滑样条是样条平滑的主流,但 MacNab (2007)<sup>[23]</sup> 认为, P 样条与平滑样条主要在大样本下可以得到较好的结果,但其在样本量相对有限时,这些方法很可能具有相当大的贝叶斯敏感性与不确定性,而此时回归 B 样条则能够更好地拟合样本数据。考虑到本研究包含的调查区域与时间段均有限,同时为保证拟合曲线的适度平滑性,这里我们以回归立方 B 样条为基函数。

$RS_{0i}(t)$ 、 $S_j(t)$  与  $RS_{ji}(t)$  分别表示为:

$$RS_{0i}(t) = \sum_{k=1}^K b_{ik} B_k(t), \quad (15)$$

$$S_j(t) = \sum_{k=1}^K a_{jk} B_k(t), \quad (16)$$

$$RS_{ji}(t) = \sum_{k=1}^k \beta_{jik} B_k(t), \quad (17)$$

其中,  $a_{jk}$  ( $j = 1, 2, k = 1, \dots, K$ ) 是固定样条系数;  $B_k$  ( $k = 1, \dots, K$ ) 是一序列 B 样条基函数,  $B_k(t)$  表示在时间点  $t$  ( $t = 1, 2, \dots, 8$ ) 处的第  $k$  个 B 样条基函数 ( $K = L + 3$ ),  $L$  表示结点数;  $b_k = (b_{1k}, \dots, b_{Nk})^T$  与  $\beta_{jk} = (\beta_{j1k}, \dots, \beta_{jNk})^T$  为随机样条系数。需要指出的是,为保证模型的可识别性, B 样条基函数  $B_k$  不包含截距项。

结点数及结点放置的位置是样条回归中的重要问题。Ruppter (2002)<sup>[24]</sup> 已经证明,当结点数充分多时,结点位置对 P 样条拟合几乎没有影响,但对于回归样条,由于没有设置惩罚项,应当谨慎选择结点。本文参照 MacNab (2007)<sup>[25]</sup> 的做法,对时间点相对较少的数据采用自适

应的结点选择方法,即首先设置一个结点,然后依次增多,并依据 DIC 准则评估结点数增加时时间平滑度与模型拟合优度的变化。

#### 2.2.4 先验分布设定

假定权重  $\log(\delta_t)$  服从  $N(0.0, 0.169)$ ; 而性别差异的时空变异  $\beta_{it}$  与随机样条系数  $\beta_{jk}$  均为 CAR 先验并假定其精度  $\tau_\beta, \tau_b$  均服从  $\text{gamma}(5, 0.0005)$ ; 固定样条系数  $a_{jk}$ 、性别特有空间变异  $b_{ji}$  与男性与女性共同的空间变异  $b_{0i}$  均假定为  $N(0, 10000)$ 。

同时,假定随机样条系数  $b_k$  为多变量条件自回归先验 (Multivariate Conditional Autoregressive Regression, MCAR), 精度矩阵为  $\Gamma_b$ , 且服从  $\text{Wishart}(R^{-1}, p)$ 。由于缺少关于精度矩阵  $\Gamma_b$  的先验信息,因此  $R$  通常被设定为单位阵 ( $K \times K$ ),  $p (\geq K)$  是形状参数,这里假定其为  $K$ , 表示无信息先验。需要说明的是,为提高模型可识别性与降低模型复杂度,这里  $b_{0i}$  与  $b_{ji}$  均被假定为非结构化的正态先验。

### 3 实证分析

#### 3.1 数据来源与样本基本特征

本研究所用数据均来源于中国健康与营养调查 (China Health and Nutrition Survey, CHNS) 项目。由于本研究旨在尝试运用拓展后的 SCM 时空模型分析高血压风险的时空演变,因此,仅选取了调查年份具有持续性的省份数据,即分别属于东部 (江苏、山东)、中部 (河南、湖北、湖南) 与西部 (广西、贵州) 的 7 个省份,时间跨度为 1989–2011 年,每次调查时间间隔不全相等。这 7 个省份的地图采用 ArcGIS10.1 裁剪并导入 Openbugs3.2.3 软件的 GeoBUGS 模块以便于建模及结果输出,SCM 模型估计与评估均借助 Openbugs3.2.3 实现。

分别观测各省份男性与女性高血压患病率随时间变化的趋势可以发现:总体上,各省男性高血压患病率基本上高于女性,且两者均具有随时间上升的趋势,并在 1997 年达到较高值。值得注意的是,贵州省的男性与女性高血压患病率在各年份基本处于最低水平,而河南省则在 2000 年之后明显高出其他省份,应当引起关注 (图 1、图 2)。

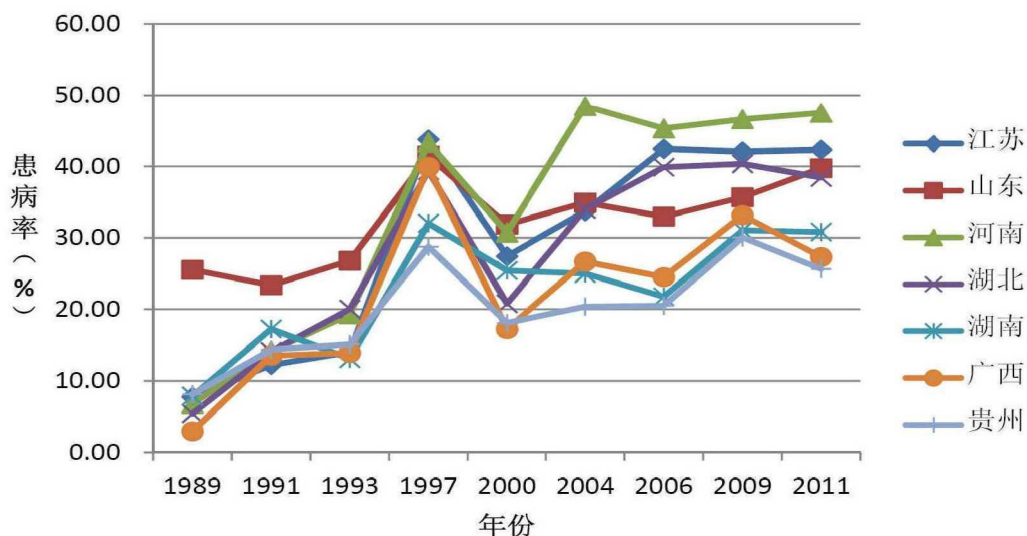


图 1 七省市男性高血压患病率随时间变化的趋势 (%)

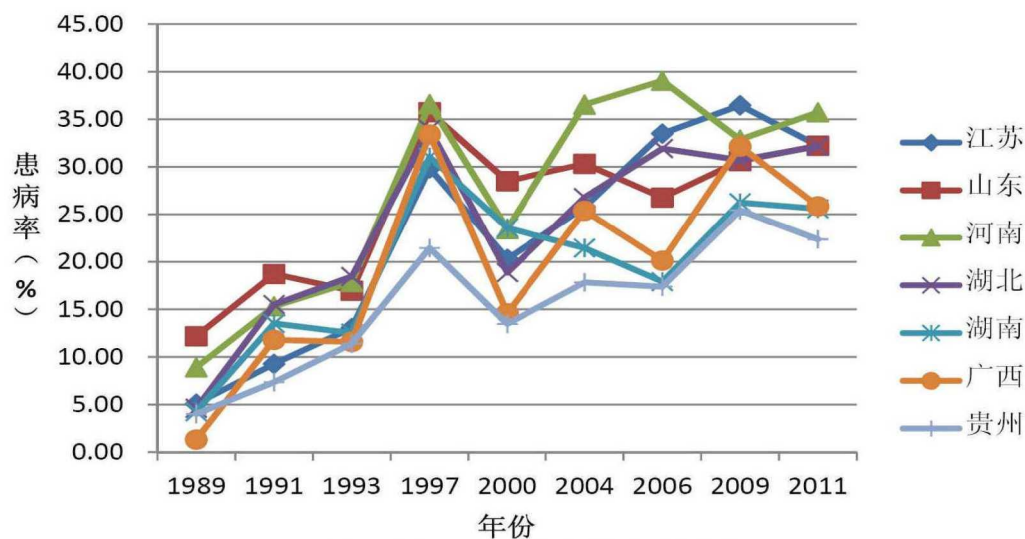


图 2 七省市女性高血压患病率随时间变化的趋势 (%)

3.2 结果分析

3.2.1 模型选择

考虑到模型成分较多可能导致模型无法识别的风险，这里结合 3.1 节描述性分析的结果首先尝试以 1997 年为结点进行建模分析 (1989 年是 CHNS 调查的第一年，样本量相对较少，可能导致区域汇总的指标不稳定，在时空分析中将其剔除)。结果发现，随机样条系数  $\beta_{jk}$  接近于 0，这意味着不存在明显的性别 - 时空交互效应，因此将  $RS_{ji}(t)$  从模型中删除。

接下来考察共同的时空主效应  $RS_{0i}(t)$  对模型拟合效果与算法运行效率的影响。结果发现，不包含  $RS_{0i}(t)$  的模型更为合理 (正的 pD, 更小的 DIC, 更少的运行时间) (表 1)，暗示不存在明显的共同的时空主效应。因此，将  $RS_{0i}(t)$  也从模型中剔除。

表 1 SCM 模型样条结点选择

模型	特点	$\overline{D}$	pD	DIC	时间 (s)
A1	Knot at 1997, with $RS_{0i}(t)$ , common interaction term	1364.0	-582.7	781.8	490
A2	Knot at 1997, with $RS_{0i}(t)$ , gender specific interaction term	1629.0	14.3	1644.0	356
A3	Knot at 1997, without $RS_{0i}(t)$ , common interaction term	1398.0	-45.4	1353.0	311
A4	Knot at 1997, without $RS_{0i}(t)$ , gender specific interaction term	1604.0	16.4	1620.0	176
B1	Knot at 2000, with $RS_{0i}(t)$ , common interaction term	1351.0	-548.0	803.2	471
B2	Knot at 2000, with $RS_{0i}(t)$ , gender specific interaction term	1612.0	-11.0	1601.0	332
B3	Knot at 2000, without $RS_{0i}(t)$ , common interaction term	1384.0	5.3	1389.0	289
B4	Knot at 2000, without $RS_{0i}(t)$ , gender specific interaction term	1589.0	-36.1	1553.0	179
C1	Knot at 1997, 2000, gender specific interaction term	1361.0	6.8	1368.0	186
C2	Knot at 1997, 2000, common interaction term	1155.0	32.7	1188	274
D1	Knot at 1997, 2004, gender specific interaction term	1529.0	-3.5	1526.0	185
D2	Knot at 1997, 2004, common interaction term	1321.0	27.9	1349	285
E1	Knot at 1997, 2000, 2004, gender specific interaction term	1338.0	19.6	1357.0	197
E2	Knot at 1997, 2000, 2004, common interaction term	1132.0	22.5	1155.0	297

最后，考察对称模型 (共同的时空交互项  $\beta_{it}$ ) 与非对称模型 (女性特有的时空交互项  $\beta_{it}$ ) 的拟合效果与算法运行效率 (表 1)。结果发现，除以 1997 年为结点的模型之外，其他模型均倾

向于对称模型, 表明存在某些共同的因素对男性与女性高血压风险产生时空交互效应。因此, 综合模型可识别性、模型拟合效果与算法运行效率等因素, 我们将模型 E2 作为最终选择。

### 3.2.2 模型诊断

模型推断采用 MCMC 法。为提高估计可靠性, 这里选择两条相互独立的马尔科夫链, 每条马氏链运行 100,000 次迭代, 并根据动态轨迹图的结果将前 50,000 次用于预迭代, 剩下的 50,000 用于推断; 同时, 为减少变量之间的相关性, 每 10 步搜集一笔资料, 从而最终的有效样本量为 10,000。

利用方差比值法统计量、动态轨迹图与自相关图等工具对模型 E2 进行收敛性诊断。结果发现, 各监测变量均具有较好的收敛性, 因此该模型能够较好地拟合男性与女性高血压风险的时空变异。

### 3.2.3 敏感性分析

考虑到本研究样本量相对有限, 先验分布的不同设定可能对模型结果产生重要影响。因此, 接下来我们针对共同成分权重  $\delta_t$  与精度参数  $\tau$  进行敏感性分析。

首先, 针对精度参数  $\tau$  考虑以下三组超先验设定: ①  $\tau \sim \text{gamma}(10.0, 1.0E-4)$ ; ② 标准差  $\sigma$  服从  $U(0, 1)$ , 并与原有的先验设定即  $\tau \sim \text{gamma}(5, 5.0E-4)$  进行比较。其次, 参照 Ancelet (2012)<sup>[26]</sup> 的做法并结合本研究特点, 考虑  $\delta_t$  的另一种先验设定, 即  $\log(\delta_t) \sim N(0.0, 0.35)$ , 用来反映两者的比值以 95% 概率介于 1/10 与 10 之间的先验信念。

对精度参数  $\tau$  的敏感性分析发现, 除  $\beta_{it}$  有轻微敏感性 (在 6% 范围内波动) 外, 其他参数, 如基线风险  $a_j$ 、固定样条系数  $a_{jk}$ 、高血压风险的 RR 值、权重  $\delta_t$  等均对  $\tau$  的先验设定不敏感。对  $\delta_t$  的敏感性分析发现, 除  $b_{0i}$  与  $a_1$  有轻微敏感性 (在 6% 范围内波动) 外, 其他参数均对  $\delta_t$  的先验设定不敏感 (详见附表 1 与附表 2, 需要可提供)。因此, 敏感性分析说明, 数据源中能够为模型中主要成分的贝叶斯学习提供足够的信息, 从而它们的后验主要由似然函数决定, 因此本文提出的 SCM 时空模型具有可靠性。

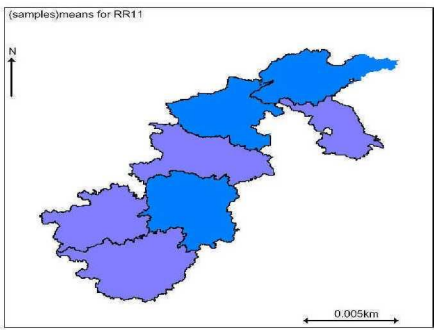
### 3.2.4 高血压相对风险 (RR) 的空间分布

图 3 为根据 SCM 时空模型拟合得到的 1991–2011 年男性与女性高血压相对风险 (RR) 后验均值的空间分布 (RR 以 0.3、0.5、0.8 与 1.0 为分割点), 图 4 为以性别特有的 ( $\exp(b_1)$ ,  $\exp(b_2)$ ) 与共同的空间变异 ( $\exp(b_0)$ ), 图 5 为 1991–2011 年时空交互项 ( $\beta_{it}$ ) 大于 1 的后验概率图, 图 6 表示男性与女性高血压风险共同变异权重  $\delta$  随时间的演变趋势, 图 7 为性别–时间交互效应 ( $S_j(t)$ ) 的变化趋势图。

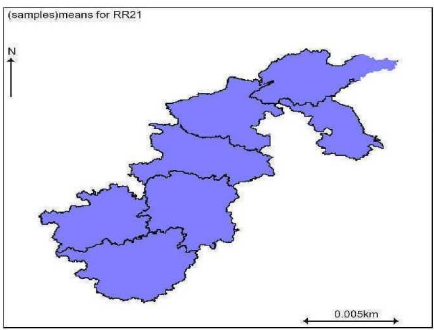
由图 3 看出, 男性与女性高血压风险的空间分布及随时间演变的规律既有相似之处也存在着明显差别。概括起来, 主要表现为以下几个特点: ① 男性与女性高血压风险的空间分布随时间演变趋势大致以 1997 年为界: 1997 年以前各省份男性与女性高血压风险均处于较低水平 (RR 后验均值不超过 0.5), 且各省份风险没有明显差别; 1997 年后, 男性与女性高血压风险上升 (尤以男性更为明显), 呈现出较为明显的北方高于南方的空间格局; ② 1997 年 (含) 之后, 河南省男性与女性高血压风险均有相对稳定的突出表现: 其中, 女性高血压 RR 后验均值除 2000 年较小之外, 其他年份均介于 0.5 与 0.8 之间; 男性高血压 RR 后验均值除 2006 与 2011 年高于 0.8 之外, 其他年份也介于 0.5 与 0.8 之间; ③ 1997 年后, 女性高血压风险, 除河南省有突出表现外, 山东省与江苏省分别在 1997 年、2006 年与 2009 年有突出表现; ④ 1997 年后的男性高血压风险, 除河南省有突出表现外, 山东省、江苏省与湖北省在大部分年份也有突出



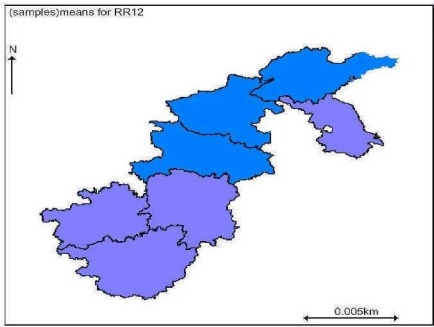
表现。



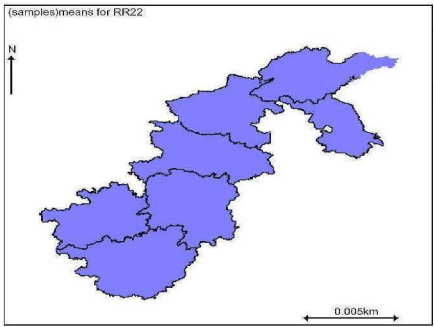
A1. 1991 RR, male



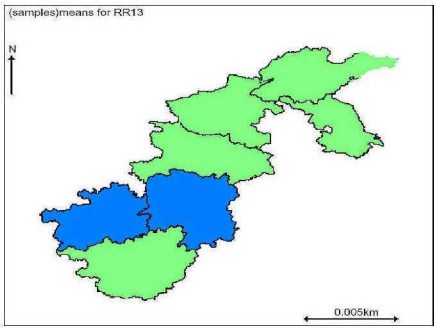
B1. 1991 RR, female



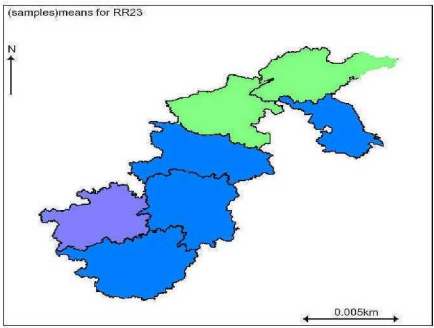
A2. 1993 RR, male



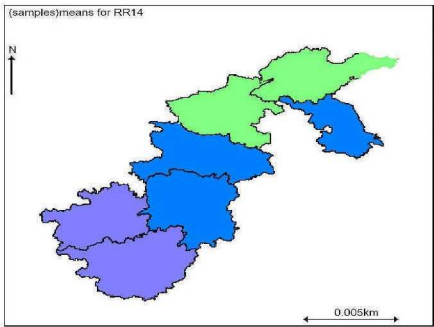
B2. 1993 RR, female



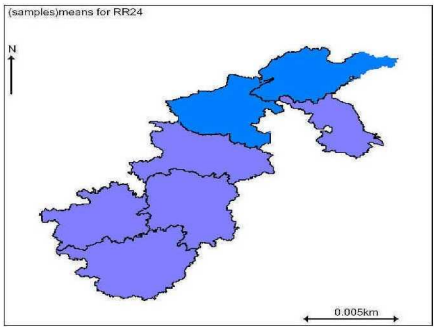
A3. 1997 RR, male



B3. 1997 RR, female



A4. 2000 RR, male



B4. 2000 RR, female



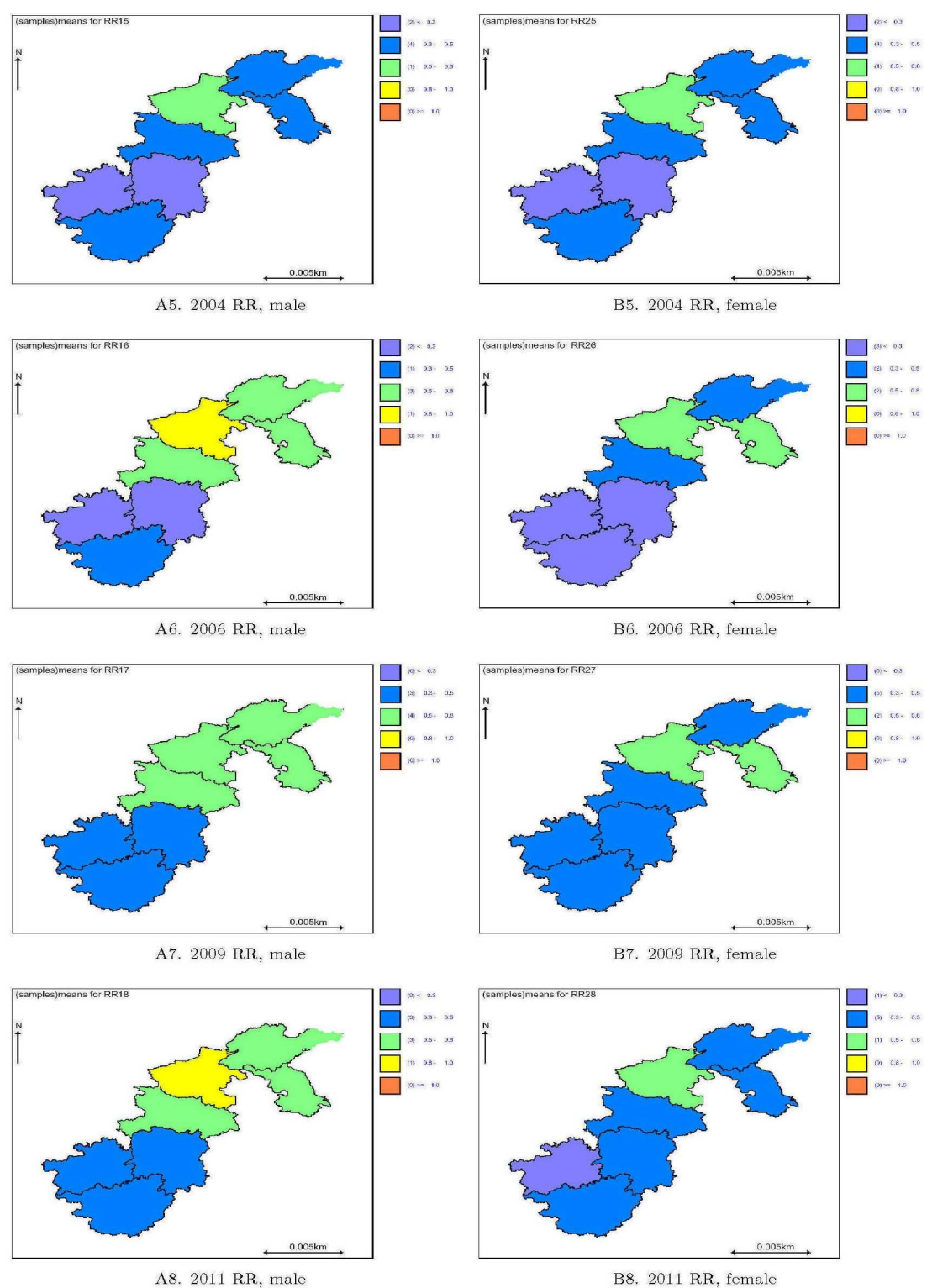


图 3 1991–2011 年男性与女性高血压风险的空间分布

由图 4 (a)–(c) 看出, 存在明显的空间主效应与性别 – 空间交互效应。其中, 河南省的空间主效应有突出表现 (图 4 (c)), 这点也可以从图 3 得到验证; 各地区男性 – 空间交互效应存在

着较为明显的差异, 尤为河南省表现最为突出 (图 4 (a)); 尽管女性高血压风险普遍低于男性, 但各地区的性别 - 空间效应相对于男性更为明显, 尤以广西表现最为突出 (图 4 (b))。总之, 图 4 (a)–(c) 进一步表明, 存在着共同的与性别特有的因素主导着男性与女性高血压风险空间分布, 由此说明采用 SCM 模型拟合具有合理性, 而共同成分权重  $\delta$  后验均值始终小于 0.35 也验证了这一点 (图 6)。

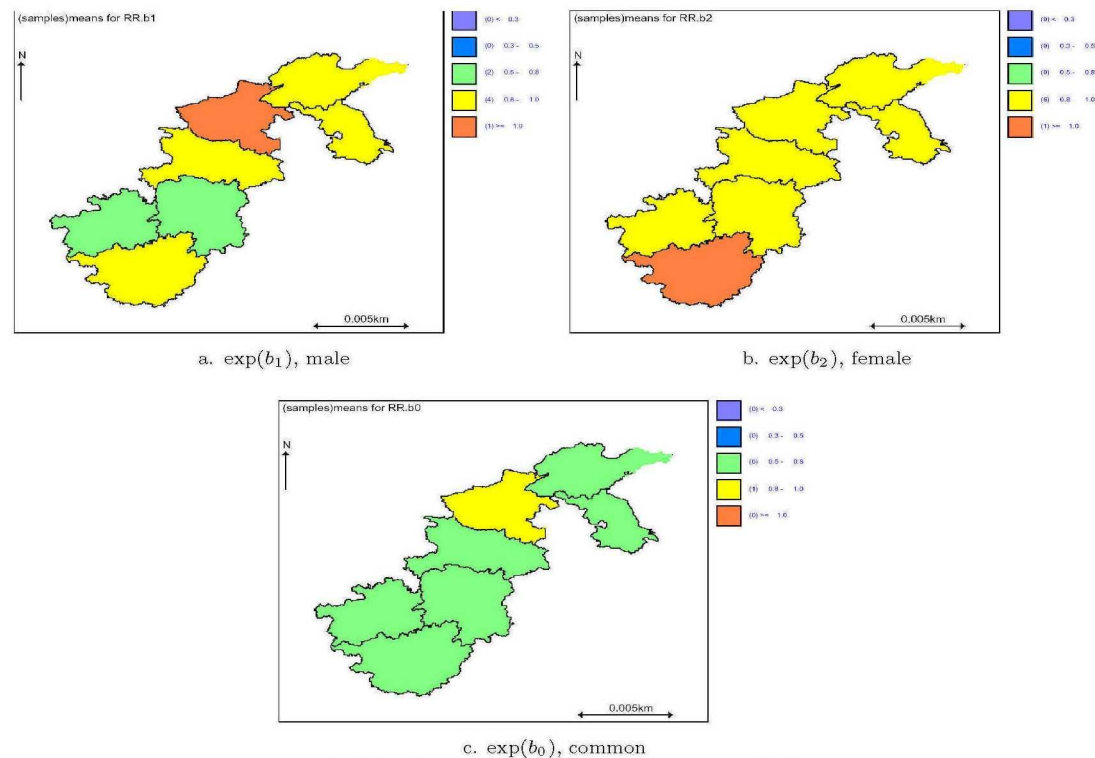


图 4 (a)–(c) 性别特有的 ( $\exp(b_1)$ ,  $\exp(b_2)$ ) 与共同的空间变异 ( $\exp(b_0)$ )

进一步, 由图 5 (a)–(h) 看出, 男性与女性高血压风险时空交互效应较为明显且随时间推移逐渐向北移动, 从而呈现出与高血压风险相似的北方高于南方的空间格局。这表明, 对于高血压而言, 在相对较长的时期内 (比如 20 年), 存在某些因素对高血压产生时空交互作用。

最后, 本研究看到了明显的性别 - 时间交互效应。其中, 男性高血压风险时间效应更为明显, 表现为 1997 年突然上升以及 2000 年之后稳步上升的趋势, 而女性高血压风险时间效应则没有突出表现 (图 7), 由此表明, 高血压风险性别差异随着时间推移有扩大趋势。

总之, 研究表明高血压风险 (尤其是男性) 具有较为明显的空间集群性与时间集群性, 即空间上相邻的省份不仅高血压风险较为相近且随时间变化的趋势也存在着相似性, 而且存在着共同的与性别特有的因素对高血压风险产生影响, 使得男性与女性高血压风险既有相似之处又存在着明显的性别差异。由此表明, 本文提出的 SCM 时空模型能够突出高血压风险空间分布的持久性, 为主导疾病风险的因子提供指示性信息。因此, 将贝叶斯样条引入 SCM 模型进行拓展后的时空模型, 能够同时利用疾病的相关性与疾病的时空信息, 为多疾病的风险预防与监测提供了更为丰富的信息。

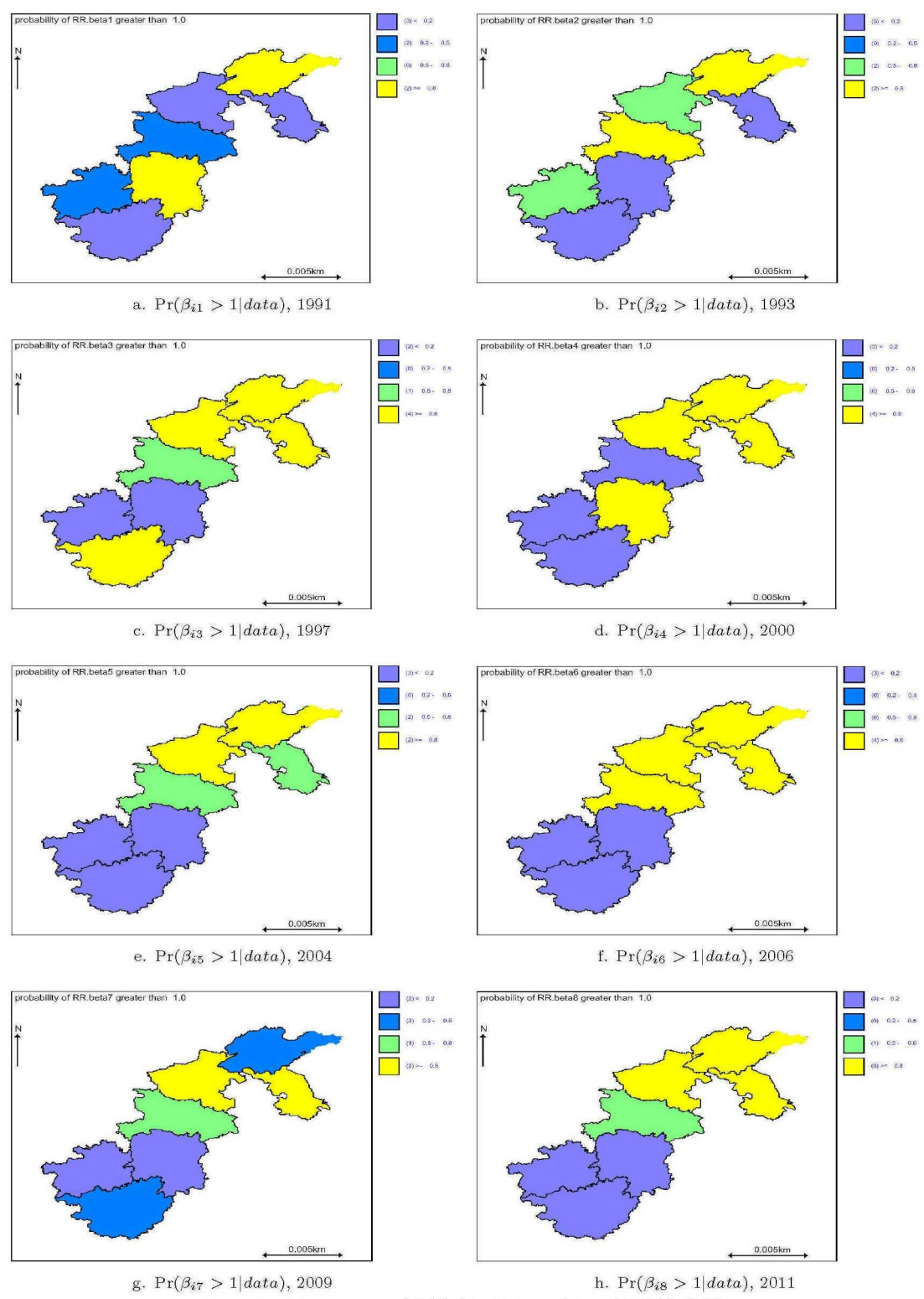


图 5 (a)–(h) 1991–2011 年时空交互项 ( $\beta_{it}$ ) 大于 1 的后验概率图

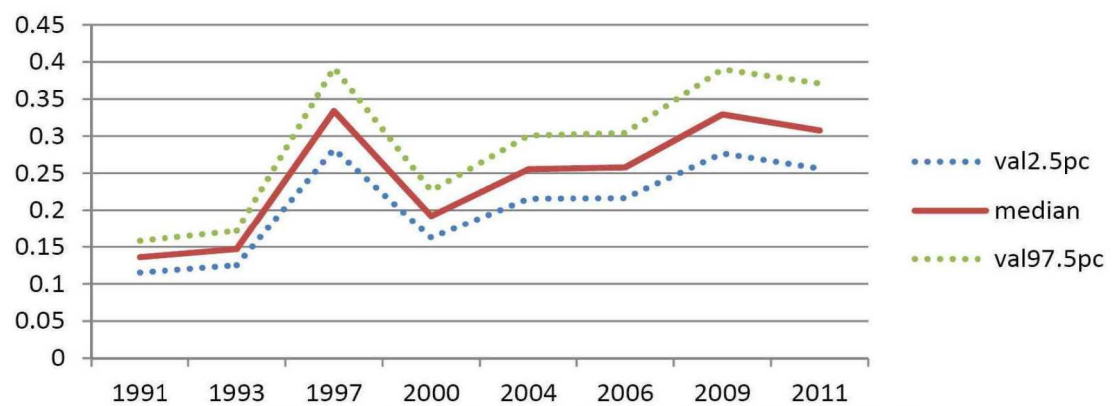


图 6 1991–2011 年共同成分的权重  $\delta$  (中位数及 95%CI) 的变化趋势

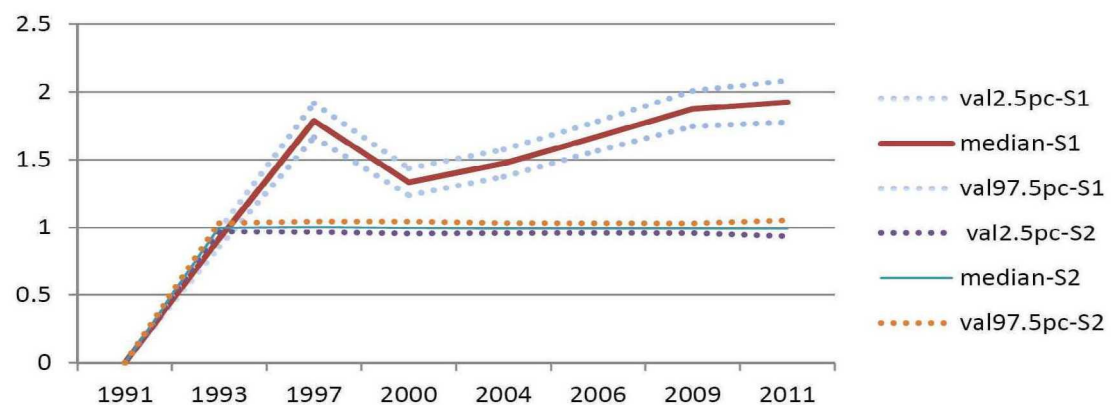


图 7 性别 – 时间交互效应 ( $S_j(t)$ ,  $j = 1$  男性) (中位数及 95%CI) 变化趋势

4 结论与政策建议

4.1 结论与不足

本研究尝试将贝叶斯样条融入贝叶斯共同成分模型，构建贝叶斯时空模型，并基于 CHNS 数据对 1991–2011 年我国七省份 (山东、江苏、河南、湖北、湖南、贵州和广西) 男性与女性高血压风险的时空变异进行了分析，为我国不同地区、不同人群的高血压风险评估提供指导信息。

研究发现，高血压风险存在性别差异且随着时间推移有扩大趋势，即存在着共同的与性别特有的因素对男性与女性高血压风险产生相应作用，表明对于本研究而言 SCM 模型的基本假定 (疾病之间具有相似性) 是合理的。

研究还发现，男性与女性高血压风险的空间分布随时间演变趋势大致以 1997 年为界：1997 年以前各省份男性与女性高血压风险均处于较低水平，1997 年后，男性与女性高血压风险上升 (尤以男性更为明显)，呈现出较为明显的北方高于南方的空间格局。值得指出的是，1997 年 (含) 之后，河南省男性与女性高血压风险均有相对稳定的突出表现，而山东省、江苏省与湖北省男性高血压风险在 1997 年后有相对明显的突出表现。因此，在高血压防控中需要特别关注这些区域。

最后，研究表明高血压风险具有相对明显的空间集群性与时间集群性，即空间上相邻的省份不仅高血压风险较为相近且随时间变化的趋势也存在着相似性。这一空间集聚性表现为高血压风险 (尤其是男性) 大致呈现出北方高于南方的空间格局，时间集群性表现为高血压风



险空间分布随时间演变规律具有相对稳定性。而时空交互效应与高血压风险的空间格局与演变规律的大致相似(尤其是男性), 则表明北方较高的高血压风险很可能与北方地区各省份相似且稳定的生活方式(喜好饮酒、膳食高盐等)有较大关联, 验证了已有研究的结论<sup>[3,5]</sup>。总之, SCM 时空分析暗示喜好饮酒、膳食高盐等生活方式的相似性很可能与高血压风险的空间集聚性与时间集群性有较大关系。

本研究的主要贡献在于: ①首次尝试将贝叶斯样条引入 SCM 模型, 构建 SCM 时空模型, 丰富 SCM 模型的理论研究。SCM 时空模型能够同时考察疾病风险的空间相关性与时间相关性, 突出疾病风险空间分布的持久性模式, 有助于识别导致疾病风险的系统性因素。②尽管现有 SCM 模型主要关注罕见疾病, 但本研究(及作者之前的研究, 见文[27])表明, SCM 模型也能够用于研究非罕见疾病风险的空间(或时空)变异, 丰富 SCM 模型的应用研究。③以性别为研究视角, 动态评估我国七个省份的高血压风险, 有助于卫生相关部门按地区、按人群制定针对性的预防与干预策略。

本研究的主要不足是: ①以省级为分析尺度, 可能导致参数的统计推断不确定性较大, 但由于调查数据的保密性等原则性要求, 无法获得尺度更细的市或县的数据; ②高血压风险的空间分布是否与饮酒、膳食高盐等指标存在统计上的关联还需通过模型加以验证, 而本研究由于篇幅的限制主要结合文献研究的结果进行分析, 充分性略显不足, 这是本研究下一步要做的工作。

总之, 本研究表明, 拓展后的 SCM 模型能够同时利用样本中疾病的相关性与时空相关性信息, 更充分地挖掘样本信息, 克服了本研究空间样本量相对不足的局限。尽管如此, 若能够获取更多的样本信息或获得尺度更细的数据(如市或县), SCM 时空模型的相对优势将会更为突出。

## 4.2 政策建议

### (一) 因时因地制宜, 制定高血压防治中长期规划

各级政府和卫生行政主管部门在高血压防治中应当因时、因地制宜, 制定高血压防治中长期规划, 并保障高血压防治的经费投入, 鼓励和支持高血压人群防治研究。在高血压防控工作中应当有所侧重, 加强高血压高发地区与高危人群的监测力度, 有效降低高血压高发地区的疾病负担。

### (二) 推广健康生活方式, 控制高血压危险因素流行

剔除遗传因素外, 高血压很大程度上是可防可控的, 其形成不是一朝一夕的, 而与长期以来的不良生活方式有重要关联, 如频繁饮酒、膳食高盐等。因此, 积极引导健康的生活方式, 推动合理、平衡的饮食方式, 提倡低脂肪、低盐饮食, 控制高血压危险因素流行是防治高血压的基本步骤。

## [ 参考文献 ]

- [1] 刘力生. 中国高血压防治指南 2010 [J]. 中华高血压杂志, 2011, (8): 701-743.
- [2] Who W H O. A global brief on Hypertension-Silent killer [R]. Global Public Health Crisis, 2013.
- [3] 罗雷, 栾荣生, 袁萍. 中国居民高血压病主要危险因素的 Meta 分析 [J]. 中华流行病学杂志, 2003, (1): 56-59.
- [4] 李新建, 徐继英, 姚海宏. 上海市居民高血压危险因素病例对照研究 [J]. 中国慢性病预防与控制, 2010, (4): 375-376.
- [5] Niskanen L, Laaksonen D E, Nyyssönen K, et al. Inflammation, abdominal obesity, and smoking as predictors of hypertension [J]. Hypertension, 2004, 44(6): 859-865.

- [6] 王丽娜, 曹丽, 张敬一, 等. 河北省成年居民高血压患病状况及相关危险因素分析 [J]. 中国慢性病预防与控制, 2008, 16(2): 125–129.
- [7] 肖静, 杨梅, 宗蕾, 等. 高血压健康危险因素的 logistic 回归和对数线性模型分析 [J]. 中国慢性病预防与控制, 2012, 20(4): 372–374.
- [8] 刘红艳, 彭辉, 刘奥博, 等. 中国人原发性高血压危险因素 Meta 分析 [J]. 中国心血管杂志, 2009, 14(3): 205–209.
- [9] Kagan A, Faibel H, Ben-Aric G, et al. Gender differences in ambulatory blood pressure monitoring profile in obese, overweight and normal subjects [J]. Journal of Human Hypertension, 2007, 21(2): 128–134.
- [10] Pechère-Bertschi A, Burnier M. Female sex hormones, salt, and blood pressure regulation [J]. American Journal of Hypertension, 2004, 17(10): 994–1001.
- [11] Sesso H D, Cook N R, Buring J E, et al. Alcohol consumption and the risk of hypertension in women and men [J]. Hypertension, 2008, 51(4): 1080–1087.
- [12] Cappuccio F P, Stranges S, Kandala N, et al. Gender-specific associations of short sleep duration with prevalent and incident hypertension the Whitehall II study [J]. Hypertension, 2007, 50(4): 693–700.
- [13] Knorr Held L, Best N G. A shared component model for detecting joint and selective clustering of two diseases [J]. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2001, 164(1): 73–85.
- [14] Dabney A R, Wakefield J C. Issues in the mapping of two diseases [J]. Statistical methods in medical research, 2005, 14(1): 83–112.
- [15] Held L, Nat á rio I, Fenton S E, et al. Towards joint disease mapping[J]. Statistical methods in medical research, 2005, 14(1): 61–82.
- [16] Richardson S, Abellan J J, Best N. Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK) [J]. Statistical Methods in Medical Research, 2006, 15(4): 385–407.
- [17] Tzala E, Best N. Bayesian latent variable modelling of multivariate spatio-temporal variation in cancer mortality [J]. Statistical Methods in Medical Research, 2008, 17(1): 97–118.
- [18] Baker J, White N, Mengersen K, Rolfe M, Morgan G G. Joint modelling of potentially avoidable hospitalisation for five diseases accounting for spatiotemporal effects: A case study in New South Wales, Australia [J/ON]. PLoS ONE, 2017, 12(8): e0183653. DOI: 10.1371/journal.pone.0183653.
- [19] Mahaki B, Mehrabi Y, Kavousi A, et al. A spatio-temporal multivariate shared component model with an application to Iran cancer data [R]. arXiv preprint arXiv: 1707.06075.2017.
- [20] 王晓荷, 田茂再. 基于分层贝叶斯时空模型的雾霾天气过程计数分析 [J]. 数理统计与管理, 2017, 36(6): 970–982.
- [21] 苏珊娜·霍夫曼 (Susanna Hoffman) (著), 陈梅 (译). 人类学视野中的灾难和文化要素的探讨 [J]. Journal of Ethnology, 2015, (4): 29–35. DOI: 10.3969/j.issn.1674-9391.2015.04.005.
- [22] Spiegelhalter D J, Best N G, Carlin B P, et al. Bayesian measures of model complexity and fit [J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2002, 64(4): 583–639.
- [23] MacNab Y C. Spline smoothing in Bayesian disease mapping [J]. Environmetrics, 2007, 18(7): 727–744.
- [24] Ruppert D. Selecting the number of knots for penalized splines [J]. Journal of Computational and Graphical Statistics, 2012, 11(4): 735–757.
- [25] MacNab Y C, Gustafson P. Regression B-spline smoothing in Bayesian disease mapping: With an application to patient safety surveillance [J]. Statistics in medicine, 2007, 26(24): 4455–4474.
- [26] Ancelet S, Abellan J J, Del Rio Vilas V J, et al. Bayesian shared spatial-component models to combine and borrow strength across sparse disease surveillance sources [J]. Biometrical Journal, 2012, 54(3): 385–404.
- [27] Li Xu, Dejian Lai, Ya Fang. Spatial analysis of gender variation in the prevalence of hypertension among the middle-aged and elderly population in Zhejiang Province, China [J]. BMC Public Health, 2016, 16: 447. DOI: 10.1186/s12889-016-3121-y.